IDENTITY & SECURITY

Take charge of your data: using Cloud DLP to de-identify and obfuscate sensitive information

Jordanna Chord
Engineer, Google Cloud

Scott Ellis
Product Manager, Google Cloud

November 20, 2018

In our previous "Taking charge of your data" post, we talked about how to gain visibility into your data using the Cloud Data Loss Prevention (DLP) API. But discovering sensitive data is just the start. In this post we'll tackle how to protect that data by incorporating data obfuscation and minimization techniques automatically into your

Find an article...

Latest stories

Products

Topics

About

RSS Feed

identification. NIST defines de-identification as the process of "remov[ing] identifying

Thus, with de-identification, if a customer gave their phone number in a chat message like:

```
My phone number is 8582394000 if you need to reach me
```

You could share the text minus the phone number—something like:

```
My phone number is [PHONE_NUMBER] if you need to reach me
```

In this way, de-identification can help reduce the risks inherent in data so that when someone is granted access to it, they are less likely to be exposed to any sensitive PII.

De-identification in Cloud DLP

There are several de-identification techniques that can help obscure sensitive information while preserving some utility. Below are a few common techniques supported in Cloud DLP.

- Replacement - Replaces each input value with a given value.
- Redaction - Redacts a value by removing it.
- Mask - Masks a string either fully or partially by replacing a given number of characters with a specified fixed character. This technique can, for example, mask everything but the last 4-digits of an account number or social security number.

Find an article...

Latest stories

Products

Topics

About

RSS Feed

ranges within which the input value falls. For example, you can bucket specific ages

inside a block of text.

Expanding on the phone number example above, let's say you wanted to replace the phone number using *pseudonymization with a format preserving token* This would look something like the following:

```
My phone number is 6070548884 if you need to reach me
```

Or, here's the pseudonymized output plus an optional prefix indicating it can be reverted:

```
My phone number is PHONE(10):6070548884 if you need to reach me
```

All of Cloud DLP's de-identification options are available through a simple REST API, as well as several client libraries in common scripting and programming languages. To get started, check out the Cloud DLP documentation.

Cloud DLP in action

Cloud DLP provides a variety of flexible and scalable tools to help you de-identify sensitive data and reduce risk in your production workloads. Here are a few examples of how we see customers using Cloud DLP.

Automated large-scale data obfuscation

Your business needs to share data with a third-party to run an analysis, but you don't

Find an article...

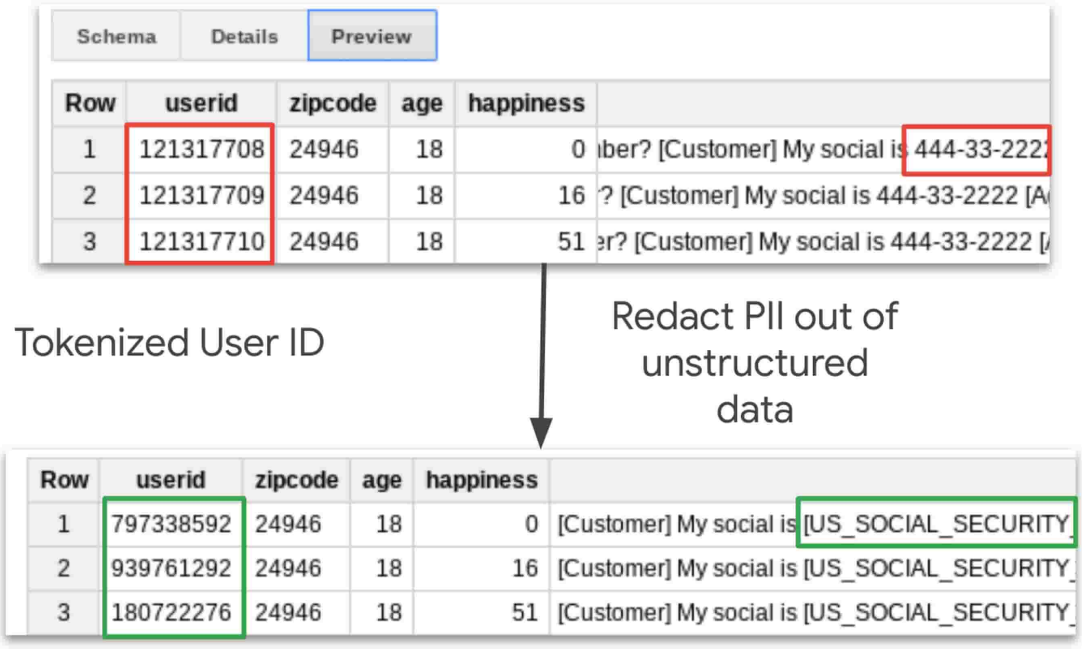
Latest stories

Products

Topics

About

RSS Feed



To get started, here is a reference pipeline that leverages Cloud DLP and Dataflow to automatically obfuscate data and ingest it to Cloud Storage or BigQuery.

Real-time data minimization

In the spirit of reducing privacy risks, you need to manage the data that you collect from

Find an article...

Latest stories

Products

Topics

About

RSS Feed

Better yet, you can call Cloud DLP's API directly to integrate with virtually any application, including Google Apigee, our API management platform, to help protect all your API endpoints inbound and outbound.

Preventing PII stored in your data from exposure is a key concern for many organizations—and not so easy to do. Cloud DLP provides powerful tools to help you protect the security and privacy of your data, via an easy-to-use and flexible API. To learn more, visit our Cloud Data Loss Prevention page for more resources on getting started.

POSTED IN: IDENTITY & SECURITY—GOOGLE CLOUD PLATFORM

RELATED ARTICLES

Using Cloud Logging as your single pane of glass

Under the hood: The security analytics that drive IAM recommendations on Google Cloud